An Ontological Approach to Efficient Text Based Search In Websites

Shah Nazir, Sara Shehzad, Neelam Mukhtar
Department of Computer Science
University of Peshawar
Pakistan
shahnzr@upesh.edu.pk, sara@upesh.edu.pk, sameen_gul@yahoo.com

ABSTRACT: *As the number of internet users and searchable content is increasing day by day, the text and documents become more complex and are becoming more difficult to find and investigate. Efficient search and retrieval of the required information is very critical for the internet users which is becoming a challenge day by day. In this paper we present an ontological model for efficient text based search functionality. The goal is to let the user explore all available options to find and retrieve the required piece of information. The ontological model is presented for a university website as a case study and some usage scenarios are given.*

Keywords: Text Processing, Ontology, Web Search, Web Retrieval, Document Processing

Received: 15 December 2013, Revised 16 January 2014, Accepted 21 January 2014

1. Introduction

With the large and growing information available on the World Wide Web (internet), the job of retrieving text and documents of interest has become more than a challenging job. Users have two ways to find and locate the information: they can do keyword based searching in a search engine, such as, Google and Yahoo so on and information users are allowed to enter some keywords and on the basis of the given keywords, the relevant documents and information is displayed, they can browse a web page selecting from hierarchy of categories till the required information is reached. The relevant sub-category provides the related information. The most simple and comprehensible way of finding documents and information through a given query is to check, that the documents holds the requested and required information.

The goal of this paper is to present a methodology for an efficient text based search. In this paper an ontological text based search approach is used for searching the required text within a website.

The paper is organized as follows. Section 2 presents literature study. Section 3 highlights the details of the proposed method. Section 4 discusses the case study. A conclusion is presented in section 5.

2. Literature study

Several diverse methodologies are proposed by the community of software engineers for the retrieval of information. D. Vallet et al. presented a model for information retrieval based on ontology. Their model improved search on large documents and with help of ontology. The improvement in this approach is with respect to keyword based search [1]. M. Wessel and R. Moller designed a framework based on description logic which also provides reasoning facilities. The basic idea is described as a framework of ontology-based query-answering system for geographical information system [2]. A.M. Figueiredo et al. suggested a framework for searching software architecture related information. An ontology-based semantic searching technique is developed for information retrieval of architecture properties. The proposed ontology in this technique facilitate very interactive atmosphere for the user to use that information [3]. J. Kang and K. Mong Sim proposed Cloudle search engine for cloud computing system. Three types of reasoning methods are determined which are similar to cloud services. The reasoning methods are concept similarity, object property similarity and data type similarity [4].

T. Heeptaisong and A.Srivihok suggest some steps to design soil ontology for soil knowledge searching. The approach searches knowledge in different available sources on the internet. This knowledge can be reused and shared. A total of 84 nodes and 83 relationships are designed for the ontology of seed soil. It is also verified by the soil expert and mentioned that it is effectively perfect [5]. J.A. Gulla et al. discussed the extraction system for learning of ontology. Then they assessed the output of ontology search driven applications. They showed that when using learned ontology, the quality is slightly affected for searching [6]. S. Gauch et al. reported to examining the communication of conceptual search with the user profiles. They showed that the created profiles reflect the interest of users and produced reasonable improvement to search results [7].

Y. Zhang et al. worked on OntoSearch approach a kind of Google ontology. This approach allows user to provide a key word for ontology based search and also check their relevance[8]. A.M. Esa et al. discussed the basic problems of searching on the web. The method emphasized on the semantic abilities for search engine using ontology. This method also semantically links the search results for the stored information extraction on ontology [9].

H. Alani et al. proposed a solution for the problems of ontology based search and finding the most suitable and for finding the best suited ontology according to user given query. Their system search for a specific topic and provides the relative terms to the given query. The method and algorithm is validated using ontology of biomedical domain repository [10]. T. Heeptaisong and A. Shivihok used ontology for soil knowledge based system. XPath algorithm and automatic weighting ontology are used for the development of the system. This approach finds different sources on internet for reuse and helps knowledge sharing. The system is compared with others which showing that it is efficient than traditional one [11].

3. Methodology

The proposed ontological method is based on specifications given in [12], and is designed using Protégé software [14]. The method consists of the following phases:

3.1 Determine scope

The first and the most important phase of designing an ontology is to determine the domain and scope of the ontology. Also, what type of questions the information in the ontology should answer. The domain can be so wide, that one may create ontology for each part of the domain so that the domain becomes quite explicit.

3.2 Identify competency question

The scope of the ontology can be defined in some competency questions. Such questions provide a simple mechanism to test the ontology during the process of development. For example, which characteristics should be considered when choosing different attributes. The following are some of the competency questions which we have suggested.

a. How the search mechanism can help users?

b. How a text can be searched in an efficient way?

c. Who can benefit from the search mechanism.

3.3 Consider reuse

Select the most appropriate and suitable component to fulfill specific needs of the user. Reuse can save time and effort [13]. The

reuse of an available ontology also gives error free design, because it is already being used and pre-tested.

3.4 Identify concept

According to the competency questions concept or class can be identified. A class is a concept in the domain. What are the terms we need to talk about and what are the different properties of these terms. More concept or class should be considered for the explicit information associated with a domain. Concepts are just like nouns.

3.5 Insert concept into the taxonomy

Concepts or classes are inserted into the domain taxonomy. In the proposed method we have defined the hierarchical relations among concept and/or classes. Classes usually constitute a taxonomic hierarchy. The different levels in the hierarchy are shown in figure 1.

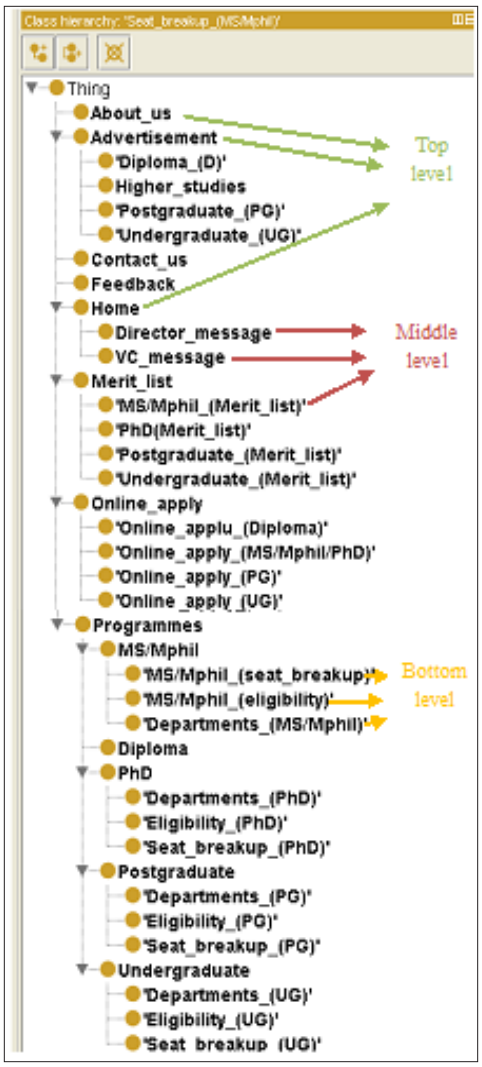

Figure 1. Proposed ontological hierarchical structure

3.6 Define properties

Class definition describes attributes of instances of class and relation to other instances. Different types of properties exist, such as simple and complex properties. Adding more properties can help to answer the competency questions.

3.7 Define constraints

Properties' constraints (facets) describe and limit the set of possible values for a slot. Common facets are string, number, Boolean, enumerated type and complex type.

3.8 Create instances

Revise the designed ontology until it is according to user requirements. Use some more concepts or classes of the domain. The term concept is mostly used than class and property rather than slot as these terms are closer to domain expert.

4. Case study

The method for efficient text based search mechanism for website is validated by using the case study of searching information in the web page of admissions section in a public sector university website. User can search information about different concepts related to the domain of admission system of the university. The information contained in the domain is shown hierarchically in figure 1. User can search any of the given information in the domain of the website of the admissions system.

The proposed search mechanism facilitates in two ways:

4.1 Generic search

This type of search mechanism is used to search text based on some keywords. The keywords are searched in all of the pages in the domain of admissions. The most relevant text of the searched query is displayed to the user.

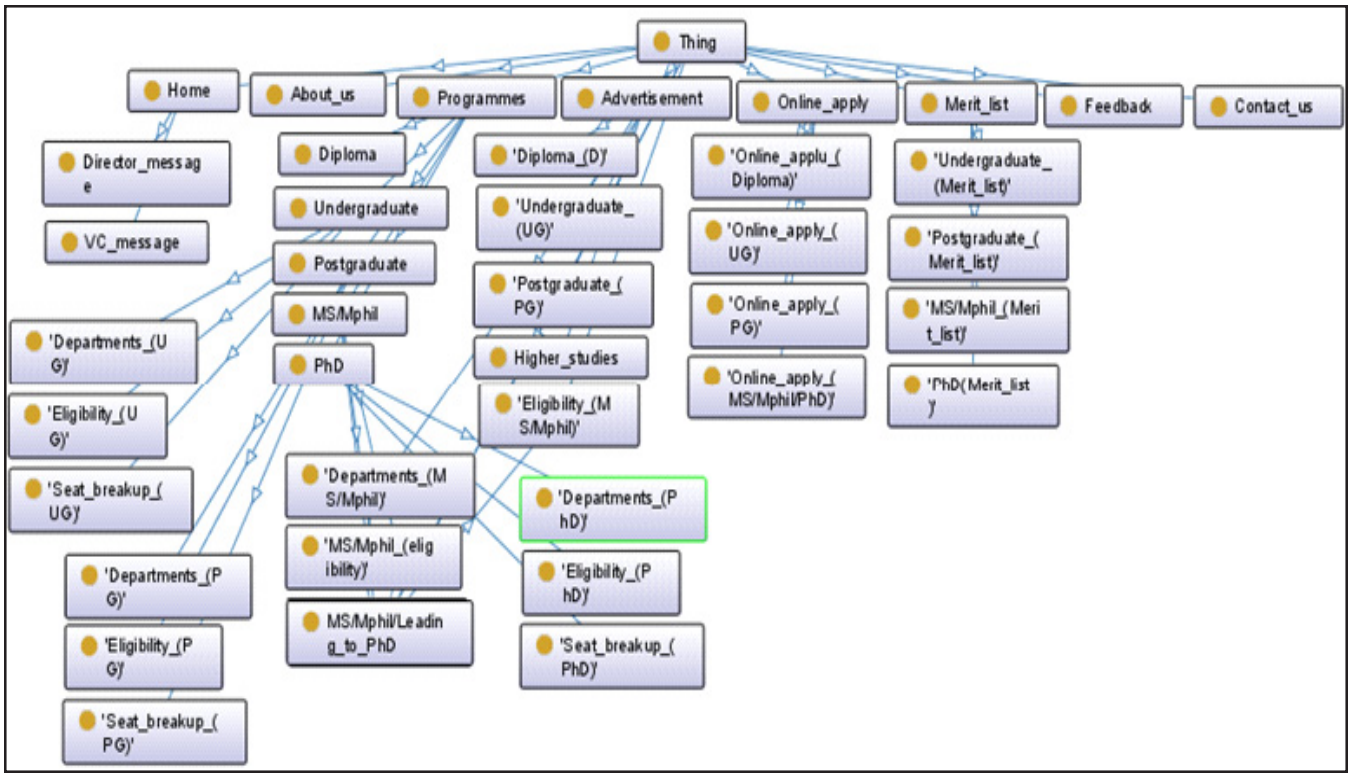

Figure 2. Proposed ontological structure

4.2 Criteria based search

In this searching mechanism a query is used for departmental criteria. If user needs to know the details of departments, in which he/she is interested to take admissions. He/she enters the department name, and the searching mechanism will display the criteria for the relevant department. For example, user needs to know the criteria of admission in the department of computer science, so the user will provide the word "computer science" to the search mechanism. As a result the relevant criteria will be displayed. Figure 2 is the structure of the proposed ontology.

5. Conclusion

The number of internet users and web pages are ultimately rising, finding a relevant document is becoming a challenging job.

Users browse related category for finding information related to a given query, most of the information that the user gets is irrelevant. The proposed methodology for efficient text based search mechanism using ontology can help users to search information related to a specific domain. The user can get the required and appropriate information quickly. The method is quite helpful in retrieving information of the domain.

References

[1] Vallet, D., Fernández, M., Castells, P (2005). An Ontology-Based Information Retrieval Model, *In:* Proc. Second European Semantic Web Conf. (ESWC '05).

[2] Wessel, M., MÃ¶ller, R (2009). Flexible software architectures for ontology-based information systems, *Journal of Applied Logic,* vol. 7, p. 75-99.

[3] Figueiredo, A. M., Reis, J. C. d., Rodrigues, M. A. (2012). Improving Access to Software Architecture Knowledge An Ontology-based Search Approach, *International Journal Multimedia and Image Processing (IJMIP),* Vol. 2, March/June.

[4] Jaeyong, K., Kwang-Mong, S (2011). Ontology and search engine for cloud computing system, *In: S*ystem Science and Engineering (ICSSE), International Conference on, p. 276-281.

[5] Heeptaisong, T., Srivihok, A (2010). Ontology Development for Searching Soil Knowledge, *In:* The 9th International Conference on e-Business (iNCEB2010).

[6] Gulla, J. A., Borch, H. O., Ingvaldsen, J. E. (2007). Ontology Learning for Search Applications, *In: On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*. vol. 4803: Springer Berlin Heidelberg, 2007, pp. 1050-1062.

[7] Gauch, S., Chaffee, J., Pretschner, A (2004). Ontology based personalized search and browsing, *Web Intelligence and Agent Systems,* 1 (3-4) 219-234.

[8] Zhang, Y., Vasconcelos, W., Sleeman, D (2004). Ontosearch: An ontology search engine, *In: P*roc.24th SGAI Int. Conf. on Innovative Techniques and Applications of Artifial Intelligence, Cambridge, UK.

[9] Esa, A. M., Taib, S. M., Hong Nguyen, T. (2010). Prototype of semantic search engine using ontology, *In: Open Systems (ICOS), IEEE Conference on*, 2010, p. 109-114.

[10] Alani, H., Noy, N. F., Shah, N., Shadbolt, N., Musen, M. A. (2007). Searching Ontologies Based on Content: Experiments in the Biomedical Domain, *In:* 4th International Conference on Knowledge Capture, K-Cap'07, 2007, p. 55-62.

[11] Heeptaisong, T., Shivihok, A (2012). Soil Knowledge-based Systems Using Ontology, *In: P*roceedings of the International MultiConference of Engineers and Computer Scientists Hong Kong.

[12] Noy, N. F., McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology," Technical Report SMI-2001-0880, Stanford Medical Informatics.

[13] Nazir, S., Khan, M. A., Anwar, S., Khan, H., Nazir, M. (2012). A Novel Fuzzy Logic Based Software Component Selection Modeling, *In:* International Conference on Information Science and Applications (ICISA) p. 1-6.

[14] http://protege.stanford.edu/